MECHANISMS SUPPORTING RECOGNITION MEMEORY DURING MUSIC LISTENING

A Thesis
Presented to
The Academic Faculty

By

Brittany Graham

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science in Psychology

Georgia Institute of Technology

December, 2011

MECHANISMS SUPPORTING RECOGNITION MEMEORY DURING MUSIC LISTENING

Approved by:

Dr. Audrey Duarte, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Christopher Hertzog
School of Psychology
Georgia Institute of Technology

Dr. Paul Verheaghen
School of Psychology
Georgia Institute of Technology

Date Approved: November 1, 2011

ACKNOWLEDGEMENTS

Special thanks to my advisor Audrey Duarte for her advice and encouragement throughout this project. I would also like to thank my thesis committee members, Christopher Hertzog and Paul Verhaeghen, for their input and recommendations, particularly on the design of this thesis. Thank you to the past and present members of the Memory and Aging Lab, without whom this project would never been completed, especially Anita Hasni, Yashu Jiang, and Rose Donahue.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

We investigated the concurrent effects of arousal and encoding specificity as related to background music on associative memory accuracy. Extant literature suggested these factors affect memory, but their combined effect in musical stimuli was not clear and may affect memory differentially for young and older adults. Specifically, we sought to determine if music can be used as a mnemonic device to overcome the associative memory deficits typically experienced by healthy older adults. We used a paired-associates memory task in which young and older adults listened to either highly or lowly arousing music or to silence while simultaneously studying same gender face-name pairs. Participants' memory was then tested for these pairs while listening to either the same or different music selections. We found that young adults' memory performance was not affected by any of the music listening conditions. Music listening, however, was detrimental for older adults. Specifically, their memory performance was worse for all music conditions, particularly if the music was highly arousing. Young adults' pattern of results was not reflected in their subjective ratings of helpfulness; they felt that all music was helpful to their performance yet there was no indication of this in the results. Older adults were more aware of the detriment of music on their performance, rating some highly arousing music as less helpful than silence. We discuss possible reasons for this pattern and conclude that these results are most consistent with the theory that older adults' failure to inhibit processing of distracting task-irrelevant information, in this case background music, contributes to their elevated memory failures.

CHAPTER 1

INTRODUCTION

1.1: Using Music to Support Cognition

Music is ubiquitous in our daily lives. We listen to music when we are driving in the car on our way to work or while we are studying for our next test. What is not clear is whether music listening actually supports cognition during these kinds of complex tasks. The goal of this project was to determine if music supports cognition, and if so, what might be some of the mechanisms responsible.

There is evidence suggesting that music facilitates cognition, specifically memory. For example, musical jingles commonly used in advertising can facilitate memory for advertising slogans (Yalch, 1991). One study that shows that music listening may improve cognition reported the so-called "Mozart Effect"(Rauscher, Shaw, & Ky, 1993). This study was widely publicized as the first causal evidence that classical music can improve cognition. Rauscher and colleagues found that when participants listened to Mozart's sonata for two pianos in D major, their scores on spatial reasoning tasks increased above those of students who did not listen to music. Despite the improvement, the gains were short lived, lasting only until the participants completed the spatial reasoning task, approximately 10 to 15 minutes after listening to the music. These effects are promising and have been replicated, e.g. Rauscher et al. (1995). In a meta-analysis, however, the cognitive enhancement gained by listening to Mozart yielded an average effect size of d=0.09 (Chabris, 1999), suggesting music does not robustly improve spatial cognition. It is important to note that the studies included in this meta-analysis primarily use spatial reasoning tasks to look at cognitive enhancement and that this type of task represents only one kind of cognition. Music may also improve recall memory. Participants' were better at recalling text when it was presented with a simple melody repeated three times (i.e. three verses of text were sung to the same melody) than when it was presented only as text or when each of the verses of text were paired with separate melodies (Wallace, 1994). Presumably, when the melody was repeated and easily learned, it became an information context that facilitates recall. When the experimenters presented the melody only once with one verse of text, however, it distracted participants from encoding the text, hindering rather than enhancing performance. Under certain conditions, music listening may have notable effects on memory and cognitive function. To test this effect, we used a recognition memory task.

Music has been used in therapeutic work with special populations. For example, in a case study of a patient with severe memory deficits due to brain damage produced by herpes simplex encephalitis, the patient recalled more titles of songs when the melody of the song was played than when simply asked to recall a song (Baur, Uttner, Ilmberger, Fesl, & Mal, 2000). Music has also been used as therapy for Alzheimer's dementia (AD) patients. When older AD patients listened to music from the 1920's and 1930's for six months in group therapy settings, they subsequently remember more autobiographical information than patients who completed puzzles and other simple recreational activities (Lord & Garner, 1993). In these studies, music provided a retrieval cue that supported accurate memory retrieval. Music listening has also been shown to increase category fluency in both health older adults as well as AD patients (Thompson, Moulin, Hayre, & Jones, 1995). In this study, when healthy older adults and AD older adult patients listened to Vivaldi, they named more examples for each of the fluency categories (i.e. animals, places, etc.) than when they did not listen to music. This effect may be the result of music providing a specific retrieval context or, according to Thompson et al., the result of some attentional enhancement. This effect was not different between healthy older adults and AD patients, suggesting playing music may enhance cognition equally for both groups. When the to-be-remembered information is embedded in the musical stimuli, there may not be a benefit of music for a healthy older adult population. In a study with AD patients, memory for children's song lyrics was enhanced when the lyrics were sung compared to simply speaking the lyrics (Simmons-Stern, Budson, & Ally, 2010). This benefit did not, however, extend to healthy older adults who remembered song lyrics equally well when sung or spoken.

Much of the work with music has been done on patient populations, and it is possible that the benefits to cognition seen in special populations treated with music therapy can be extended to healthy older adults with normative cognitive decline. The specific aspect of cognition we were interested in for the present study was episodic memory due to the well-known and widespread deficits in episodic memory accuracy in healthy aging. Remembering an episode consists of successfully encoding and retrieving both the single units of information as well as the associations between these units (Naveh-Benjamin, 2000). For example, correctly remembering a conversation includes remembering what was said, to whom one was speaking, where the conversation took place, etc. Naveh-Benjamin (2000) provides one explanation of older adults' commonly observed deficits in remembering details of a previously experienced event; the so-called "associative deficit hypothesis". He argues that older adults are capable of encoding and later remembering the single units of an episode, e.g. what information was shared during a conversation, but have difficulty binding these units of information to each other, e.g. remembering a specific conversation shared with a specific person. The associative deficit hypothesis of aging has been assessed most often with the paired-associates learning task. In this task, two items are presented at study. During test, pairs are presented either intact, exactly as presented during encoding, or rearranged with other studied units. Sometimes new items may be presented with studied items. Participants are asked to determine whether the test pairs are intact, rearranged or contain a new item. The inclusion of unstudied, new items allows for the measurement of both item-level recognition as well as associative episodic recognition. In a recent meta-analysis, Old and Naveh-Benjamin (2008) showed that older adults consistently perform poorer than young adults on tests of memory for associative information compared to memory for item level information. This evidence and similar findings (Naveh-Benjamin, Guez, Kilb, & Reedy, 2004; Naveh-Benjamin, Hussain, Guez, & Bar-On, 2003) suggest that healthy aging, even in the absence of dementia, impairs memory for episodic associations, leaving item recognition more or less intact.

No studies have investigated whether music might facilitate associative memory accuracy in old or young adults. We discuss the potential mechanisms that may support the mnemonic benefit of music to support both young and older adults' memory. These mechanisms are not mutually exclusive, and we considered the unique benefit of each, as well as their cumulative benefit, in the experimental paradigm presented in this paper.

1.2: Research with Music as an Arousing Stimulus

One prominent hypothesis in the field of music research suggests that listening to music increases physical arousal and attention, thereby improving cognition. The time point at which arousal must be induced to enhance cognition, however, is not clear as various studies show effects of arousal at different time points during cognitive task performance. Roth and Smith (2008) showed, for example, that when participants listened to either music or general traffic noises prior to testing, they performed better on a portion of the GRE than participants who listened to silence. This suggests that there are potentially many different sounds, including music that can increase arousal to improve cognition. In another study, recall on a verbal memory task was best when arousal during encoding was high but the level of arousal during retrieval did not affect recall (Standing, Bobbitt, Boisvert, Dayholos, & Gagnon, 2008). These studies suggest that pre-test arousal or arousal during encoding as induced by music may benefit various aspects of cognition, including episodic memory.

Post-learning arousal may also benefit memory by facilitating memory consolidation. Liu, Graham, and Zorawski (2008) found that post-learning arousal, induced by arousing video clips presented after encoding, enhanced recall memory for emotional, but not neutral, pictures after a one week delay. Interestingly, the valence of the arousal, either positive or negative, did not differentially affect recall performance. That is, participants were better at recalling emotional stimuli when they were negatively or positively aroused immediately following learning, but not when they watched a neutral film, suggesting it is not the valence of arousal that is important, but simply that the participant was aroused at all. More relevant to the current study, Greene, Bahri, and Soto (2010) also induced an arousing state following study of neutral abstract shapes, but used music to do so rather than a video clip. States of arousal, crossed with emotional valence, were induced in participants by playing a selected piece of music during a rest period between study and test phases of a memory task in which participants were asked to recognize the previously studied abstract shapes. The authors found that recognition memory was best when arousal was high and a positive mood was induced and also when arousal was low and a negative mood was induced, with no differences in memory performance between these two conditions. Without making inappropriate inferences from the results, suffice to say this study shows the interaction between emotion and arousal is complex and, based on this study, somewhat hard to explain. One possible explanation for these results is that the music stimuli were pre-selected by the experimenters for arousal and valence, and the participants may not have found the music arousing in the same way as the experimenters. To avoid this possible confound, we used music rated by groups of young and older adult pilot participants so as to best select music for each age group based on that age group's own ratings rather than on our expectations for arousal induction.

The above studies show that there are multiple phases of memory, primarily encoding and consolidation, at which heightened arousal may facilitate memory. If we could use music to increase arousal, we should see concurrent increases in memory performance, possibly by increasing attention to the cognitive task. Alternatively, it is possible that playing music may distract participants during the memory task. We might therefore see that the arousal level of music negatively affects memory performance and that participants show memory impairment while listening to highly arousing music.

1.3: Encoding-Retrieval Context Specificity

One non-mutually exclusive hypothesis that may explain the effects of music on cognition is the encoding specificity hypothesis outlined by Tulving and Thomson

(1971). They suggested that memory may be facilitated if the retrieval conditions are consistent with the encoding context. They argued that retrieval cues are most effective in triggering successful recovery when they induce the same kinds of processing (i.e. context) that were engaged during encoding (Thomson & Tulving, 1970). Conversely, by utilizing cues that change the context between study and test, subsequent memory performance may be impaired. For example, in word recognition tasks, recognition is impaired when the associative context of the to-be-remembered word differs between study and test, that is, when the to-be-remembered word is paired with one word at study and a different word or presented singly during test (Dalton, 1993; Tulving & Thomson, 1971). In paired-associates tasks involving faces, when the context for remembering faces changes from study to test, be it another face paired with target one, or a descriptive phrase about the face instead of the face, or the spatial environment in which the faces are learned, participants have more difficulty remembering the faces (Watkins, Ho, & Tulving, 1976). This suggests that that if we were to maintain the associative context between study and test, we should see more successful recognition memory performance for the face pairs.

There is some evidence to suggest that we can assess the role of encoding specificity as applied to music. Standing et al. (2008) found that recall on a paired-associates memory task for words was better when the same piece of background music was played during study and test, compared to when different musical pieces within the same genre (classical) were played during study and test. In order to test a fine-grained idea of this encoding specificity hypothesis, we will manipulate the song played at study and test to determine if, similar to Standing et al., the music played between study and test must be identical to reinstate the associative context at test or if it can merely be similar between the two i.e. two musical pieces from the same genre.

1.4: The Current Study

The present study sought to examine the use of music as a mnemonic device for associative recognition for young and older adults, and to determine which mechanisms might support this effect. Specifically, we manipulated the arousal and song selection (i.e. encoding specificity) of the music played during study and test in a paired-associates recognition memory task to determine how these factors influence memory performance. The paired-associates task measures recognition memory for faces paired with names and is a task with high ecological validity. We manipulated arousal by including both high and low arousal music. We also manipulated context specificity by playing music that was exactly the same between study and test (e.g. the same high arousal rock song) or different between study and test but within the same genre (e.g. two different high arousal rock songs). Valence and music genre were not manipulated.

Prior to this study, there was not enough background literature directly relating to this type of experiment to determine whether a memory enhancing effect of music might be the result of arousal or context specificity, or some combination thereof, and if these mechanisms would interact differently for young and older adults. Based on the previous evidence, however, we developed some reasonable hypotheses:

1) We predicted that memory performance in the high arousal conditions will be greater than performance in the low arousal conditions based on evidence reviewed earlier suggesting high arousal can enhance memory performance (Liu, et al., 2008; Roth & Smith, 2008; Standing, et al., 2008).

2) We predicted context specificity effects for memory accuracy. There is some evidence to suggest that the more different the retrieval context is from the original encoding context, the worse memory performance is (Standing, et al., 2008; Thomson & Tulving, 1970; Tulving & Thomson, 1971), suggesting that the more overlap between the original encoding context and the retrieval context, the more successful retrieval will be. Specifically, we predicted that associative memory will be better when the context is exactly the same (i.e. the same song is played at study and test) than when the context is different even though the genre is the same (i.e. different songs of the same genre are played at study and test).

There was also the possibility that playing any music at all will distract participants' attention away from the memory task. We would therefore see that memory performance in any music listening condition would be impaired relative to silence. If music did indeed distract participants, we could expect that older adults' performance would suffer more. Based on an inhibition theory (Hasher & Zacks, 1988) of aging, we would expect that if music is distracting, older adults would have a more difficult time than young adults inhibiting processing the background music, and we would see a concurrent drop in performance.

CHAPTER 2

METHODS

2.1: Participants

We recruited a total of 117 participants, 57 young (aged 18-30 years) and 60 older (aged 60-75 years) for participation in this study. One young and 12 older adults were excluded because they failed the screening for amusia, a music perception disorder. An additional 3 young and 3 older adults were excluded because they choose to drop out of the experiment, failed a neuropsychological screening, or had incomplete data, giving us a sample size of 53 young and 45 older adults. After analyzing performance data, we decided to reinstate into the sample 5 older adults participants who failed the initial amusia screening. There is no normative data for the amusia screening for older adults and given our high proportion of failure (i.e. 12 of 60 older adults failed), it is reasonable to assume the test may inappropriately screen out more older adults than young adults because it relies on perceptual abilities and working memory capacity that decline even in normally aging older adults. The conditions for reinstatement were as follows: their amusia screening scores were within one standard deviation of the mean for the age group for two of the three screening scores, and their performance data was within one standard deviation of the mean for four of the six memory task conditions. This gave us a final sample of 53 young and 50 older adults upon whom our analysis was based. Furthermore, a comparison of the 5 older adults who passed these criteria with a random 5 older adults who passed the amusia screening did not reveal any differences in associative recognition accuracy for any condition [$F(1,8)<1$].

Young adults were recruited from psychology courses at Georgia Institute of Technology and were given 2 hours of extra credit in their courses as compensation. Older adults were recruited from metropolitan Atlanta and were compensated $10 for each hour of participation plus an additional $5 for travel, for a total of $25. All participants were native English speakers, with no self-reported psychiatric or neurological disorders, vascular disease, current psychoactive drug use, or hearing problems. All participants signed an informed consent form approved by the Georgia Institute of Technology Institutional Review Board informing them of all pertinent study-relevant information and their rights as a participant.

2.2: Neuropsychological Assessment

We administered a standard battery of neuropsychological tests during the study. Participants completed this battery of memory and executive functioning tests so as to ensure no differences in performance due to cognitive impairments other than normal age-related cognitive decline. These tests were for screening purposes only; any participant whose score fell below two standard deviations below their age-adjusted mean score given their level of education was removed from the sample. Participants completed several subtests from the Memory Assessment Scale battery (Williams, 1991) to screen for general memory impairments including the digit span forward and backward task, and list learning task. Participants also completed the face-name paired recognition task, as well as the face-name delayed recognition task from this battery as these tests are highly similar to our experimental task. Additionally, participants completed Trail Making tests A and B (Reitan & Wolfson, 1985) as a measure of speed and attention, the Controlled Oral Word Association test (“FAS”) (Benton, Hamsher, & Sivan, 1983), a measure of verbal fluency, and an online variation of the Corsi block tapping task (Corsi, 1972), a span measure of spatial working memory. Participants were also screened for amusia, a music perception disorder, using an online version of the Montreal Battery of Evaluation of Amusia (Peretz, Champod, & Hyde, 2003).

2.3: Materials

Task-related stimuli included 156 unfamiliar male and 156 unfamiliar female faces paired with 156 male and 156 female names from the FERET database for a total of 312 face-name pairs. Face and name stimuli included face-name pairs of the same gender with an equal number of young and older adults faces, and were presented in black and white on a standard-resolution computer screen.

Music stimuli were selected from experimenters' personal collections, as well as from Jamendo.org, a free online music-sharing website. Effort was made to select music that would be unfamiliar to the majority of participants (e.g. not popular music, well-known classical music, etc.) out of concern that familiar songs may aid memory performance above and beyond planned condition manipulation (Purnell-Webb & Speelman, 2008). Additionally, only instrumental music was selected because verbal material in music may impair memory performance for verbal information via competition for phonological processing. Salame and Baddeley (1989) found that multiple kinds of music with vocal accompaniment disrupted high level cognitive tasks, especially when there was verbal information presented either verbally or visually, similar to how we will be presenting names in the face-name paired-associate task. All sounds files were normalized for volume at the -0.0 decibel level in order to play at the same volume, and were presented in uncompressed .wav files for best sound quality and volume. Ideal volume was set for each participant prior to the beginning of the experiment based on hearing comfort in order to ensure that older adults could accurately hear the music played. All music and sound throughout the experiment played at the same volume.

2.4: Pilot Study

Prior to the experimental study, we conducted a pilot study to ascertain the best songs to use for the experiment. 30 young and 15 older adults total participated in this pilot study. Pilot participants were recruited similarly to our experimental participants, were similar in demographic variables, and were screened for amusia and depression. We excluded two young adults from this sample; one participant’s data was lost due to computer malfunction and one participant did not pass the amusia screening. We excluded two older adults from this sample because they did not pass the amusia screening. We had a final sample size of 28 young and 13 older adults. Pilot participants received similar compensation to experimental participants and signed the same informed consent form.

We pre-selected 182 songs to be rated in the pilot study from four different music genres (jazz, rock/blues, classical, and electronica) in order to appeal to a wide variety of musical tastes. Participants listened to 15-second clips of each song, presented via computer using E-Prime. Participants rated songs for both arousal and valence based on a modified version of the Likert scales used in the International Affective Picture System rating study (Lang, Bradley, & Cuthbert, 2008). Ratings for valence were done using a -2 to +2 scale, with -2 representing very negative emotion, +2 representing very positive emotion, and 0 representing neutral or no emotion. Ratings for arousal were done using a

+1 to +5 scale, with +1 representing very little or not at all arousing and +5 representing extremely arousing. Participants rated all music for either the arousal or valence dimension first, then, in a separate block, listened to all music clips again to complete the second dimension rating, so as not to confound the two ratings. The order of dimension ratings was be counterbalanced across participants. Participants' instructions for arousal ratings were as follows: "You will rate each music clip on a scale of 1 to 5, with 1 being not very arousing and 5 being very arousing. An arousal rating of 1 would indicate the music clip makes you feel relaxed, calm, sluggish, dull, sleepy, or unaroused. An arousal rating of 5 would indicate the music clip makes you feel stimulated, excited, frenzied, jittery, or wide-awake." While we did not manipulate valence in this study, we asked participants to rate valence in order to facilitate selecting only positive valence songs as stimuli for the experimental study. Participants' instructions for valence ratings were as follows: "For the ratings of emotion, you will rate each music clip on a scale of -2 to +2, with -2 being negative and +2 being positive. An emotion rating of -2 would indicate the music clip makes you feel unhappy, annoyed, unsatisfied, melancholic, or despaired. An emotion rating of +2 would indicate the music clip makes you feel happy, pleased, satisfied, contented, or hopeful."

In order to facilitate a reasonable study session duration, and to reduce participant fatigue, pilot participants were randomly assigned to one of two groups, each of which rated one half of the total song sample, i.e. 91 songs per group, with equal numbers of each genre rated by each group.

2.5: Task Design

Descriptions of the experimental task conditions can be found in Table 1.

Table 1. Experimental conditions

Condition Number	Arousal	Same song during study and test?
1	Silence– Control 1	N/A
2	Musical rain – Control 2	Yes – musical rain
3	High	Yes
4	High	No
5	Low	Yes
6	Low	No

Arousal conditions included music that was rated by the pilot group as either high or low arousal. Within each of the arousal conditions there were 2 context manipulations wherein we manipulated the music played at study and at test: the same song between study and test or different songs between study and test but within the same genre. The experimental conditions with corresponding acronyms, therefore, were as follows: High arousal with the same song played during study and test (HAss), High arousal with different songs played during study and test (HAds), Low arousal with the same song played during study and test (LAss), and Low arousal with different songs played during study and test (LAds). We also included two control conditions. In one condition (silence), participants listened to nothing during both study and test. In another control condition (musical rain [MR]), participants listened to a non-musical sound called “musical rain” to elicit low levels of neutral arousal in order to include a low arousal control condition that is not musical. The musical rain sound was developed at the Centre for the Neural Basis of Hearing at the University of Cambridge, UK, the procedure for which is described by Uppenkamp, Johnsrude, Norris, Marslen-Wilson, and Patternson (2006). Musical rain is generated to be similar to a randomized vowel sound sequence in which pitch and frequency information are removed. The result is a track of sounds that do not resemble speech in pattern or frequency and do not elicit the percept of speech from participants. Participants completed two blocks of each experimental and control condition with 24 trials in each block, giving us 48 trials per condition. The order of the conditions was counterbalanced across participants.

We used a face-name paired-associates task to test associative memory performance. Both the face and the name were of the same gender. The experiment was presented in Eprime 2.0 on a Dell desktop computer. Figure 1 shows a diagram of the task.

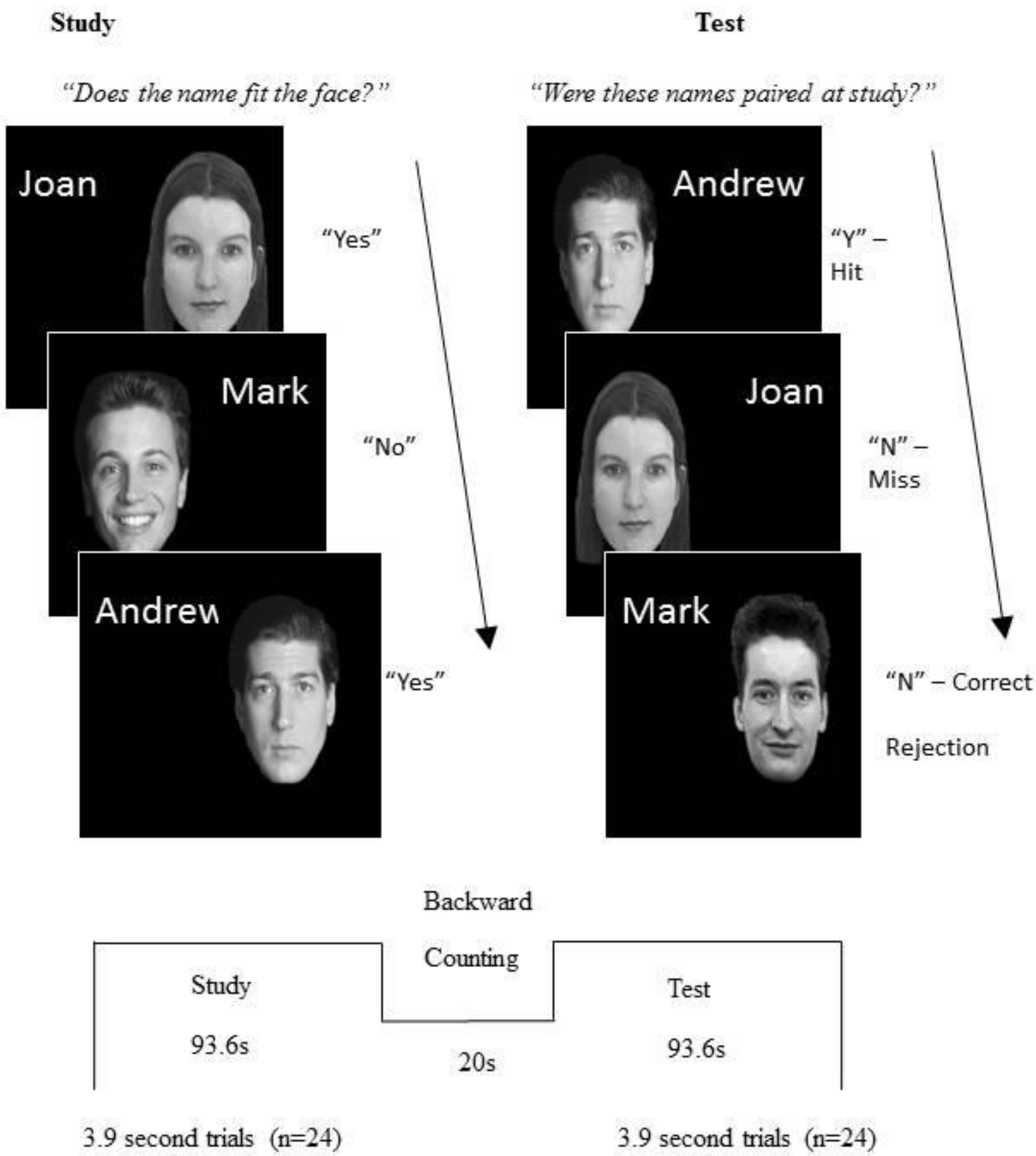

Figure 1. Diagram of experimental task design

There were 2 blocks of each of the 6 conditions, for a total of 12 study-test blocks. In each block there were 24 trials, giving us a total of 48 trials for each condition. Participants completed the 12 study-test blocks in a random order (i.e. study-test HAss, study-test LAds, etc.). During the study phase, participants studied 24 face-name pairs for 3.9 seconds each. In order to ensure successful attention to and processing of face-name pairs, participants answered the orienting question “Does this name fit/suit the face?” for each pair. After studying the 24 pairs, participants completed a backward counting task for 20 seconds to prevent them from rehearsing face-name pairs. During the test phase, participants viewed 24 face-names pairs once again and respond to the question “Were these paired during study?” During test, the positions of the faces and names were swapped (i.e. face on left at study and on right at test). This change of the items of each pair was meant to prevent unitization of the pair and the possibility that associative memory judgments would be familiarity-based (Diana, Yonelinas, & Ranganath, 2008). 16 face-name pairs were tested intact, with the identical name and face pairing and 8 pairs were rearranged. The order of the pairs was randomized at test. No new items were presented. After each study-test block, we asked participants for a rating of how helpful or distracting they found the music while they were studying and being tested. Participants responded to the helpfulness question with a 1 to 5 Likert-type scale, with 1 representing very unhelpful, 5 representing very helpful, and 3 representing neither helpful nor unhelpful. This scale was reversed for the distracting question, such that 5 represented the extreme positive, i.e. 1 meant not at all distracting, 5 represented very distracting. Between study-test blocks, participants heard neutral white noise to eliminate any residual arousing effects of the music.

CHAPTER 3

RESULTS

3.1: Pilot Study

For the young and older adult samples, we calculated separate means and standard deviations for ratings of arousal and valence for each song. We wanted to select songs within a certain range of arousal and valence ratings for young and older adults separately, and given the disparity of ratings on individual songs between the age groups, we selected different songs for the age groups. In order to determine which songs to use for the experiment, we set specific selection criteria. To be classified as high arousal, songs must have mean arousal ratings across participants greater than 3.5. To be classified as low arousal, songs must have mean arousal ratings across participants less than 3. We selected only positive valence songs for both the high and low arousal conditions because we wanted songs that were consistent in their emotional quality and for their ecological validity, our rationale being that people select positive music to listen to more often than negative music. To be classified as positive, songs had to have mean valence ratings across participants greater than 0.5. Songs were selected for use only if the standard deviations of the means were less than 1.We selected a total of 6 songs that were rated as high arousal and 6 songs that rated as low arousal songs for use in the experimental task. In addition, we used two unrated tracks of computer generated musical rain for use in the two associated control conditions. Table 2 includes mean arousal and valence ratings, as well as standard deviations for each age group for each of the songs selected for use in the study. We performed an independent samples t-test on mean ratings of arousal and valence for the songs selected. For both low and high arousal songs, young and older adults' mean arousal ratings did not differ [$t(10) < 1$], nor did they differ for valence ratings [$t(10) < 1$]. Of the four genres of music we piloted, we selected only rock songs to be played in the experiment on the basis that only this genre produced enough songs that satisfied our cutoff criteria allowing us to to select 6 high and 6 low arousal songs for each age group.

Table 2. Descriptive statistics for songs selected as musical stimuli based on pilot study of young and older adults

Musical artist, song title	Arousal Category	Mean Arousal Rating (SD)	Mean Valence Rating (SD)
Young Adults			
Pele, "Mind of Minolta"	High	3.63 (0.885)	1.06 (0.854)
John Mayall and the Bluesbreakers, "Hideaway"	High	3.75 (0.754)	1.33 (0.778)
John Mayall and the Bluesbreakers, "Curly"	High	3.92 (0.669)	1.00 (0.953)
Eric Johnson, "Cliffs of Dover"	High	4.25 (0.754)	1.42 (0.793)
Stevie Ray Vaughn, "Scuttle Buttin'"	High	4.58 (0.515)	1.58 (0.669)
Steve Vai, "Jibbom"	High	4.83 (0.389)	0.75 (1.088)
Brooks Williams, "O Leaozinho"	Low	2.19 (0.981)	1 (1.265)
Jefferson Airplane, "Embryonic Journey"	Low	2.42 (0.793)	1 (0.853)
Incredible Moses Leroy, "Roscoe"	Low	2.63 (0.957)	0.69 (1.138)
Umphrey's McGee, "Nemo"	Low	2.75 (0.856)	0.63 (1.258)
Eric Clapton, "Signe"	Low	2.75 (1.065)	1.56 (0.512)
Buddy Guy, "Just Teasin'"	Low	2.81 (0.834)	0.56 (0.814)

Table 2 (continued).

Older Adults			
Eric Johnson, “Cliffs of Dover”	High	3.67 (0.816)	1.17 (0.408)
John Mayall and the Bluesbreakers, “Steppin’ Out”	High	3.83 (0.983)	1.00 (0.00)
The Ventures, “Slaughter on Tenth Avenue”	High	4.33 (0.816)	1.00 (0.894)
Stevie Ray Vaughn, “Scuttle Buttin’”	High	4.50 (0.837)	1.17 (0.753)
Umphrey’s McGee, “The Fuzz”	High	4.50 (0.548)	1.00 (0.894)
John Mayall and the Bluesbreakers, “Hideaway”	High	4.83 (0.408)	1.67 (0.816)
Pele, “Nighttime Stomach”	Low	2.33 (0.516)	1.00 (0.894)
Incredible Moses Leroy, “Roscoe”	Low	2.50 (0.837)	0.83 (0.753)
Steve Vai, “Tender Surrender”	Low	2.50 (0.548)	0.67 (0.816)
Brooks Williams, “O Leaozinho”	Low	2.67 (1.033)	1.50 (0.837)
Jeff Beck, “Serene”	Low	2.67 (0.816)	0.67 (0.816)
Rush, “Hope”	Low	2.83 (0.753)	0.83 (0.983)

3.2: Individual Differences Results

Table 3 shows mean demographic information for young and older adults. There was no significant difference between age groups on years of education [$t(101)=-1.650$, $p=0.102$].

Table 3. Demographic information for young and older adults.

	Young	Older
Mean Age	20.92	66.30
Mean Level of Education (Years)	15.17	16
Gender (female/male)	27/26	28/22

Table 4 shows the complete mean information and age differences between young and older adults on neuropsychological screening measures as well as results from main effects of age reported from a one-way between subjects ANOVA with age group (2 levels) as the independent variable, including all of our neuropsychological measures as dependent variables. Young adults performed significant higher on the Differential Melodies and Incongruent Pause subtests of the amusia screening, and had a higher composite score than older adults. Young adults also completed significantly longer spans on the Verbal Span Forward and Corsi Block Tapping tasks, were significantly faster on the Trail Making Tests A and B.

Table 4. Young and older adult means and age group differences on neuropsychological screening measures.

	Young	Older	Main Effect of Age, *F*(1, 101)
Amusia – Differential Melodies	86.23 (8.704)	80.44 (9.702)	10.174**
Amusia – Incongruent Pause	86.40 (8.900)	81.00 (7.401)	11.123***
Amusia – Out of Tune Note	86.94 (12.141)	82.34 (13.569)	3.300, *p*=0.072

Table 4 (continued).

Amusia Composite	86.75 (6.725)	81.26 (7.551)	15.248***
Verbal Span Forward	7.35 (1.266)	6.54 (1.398)	9.115**
Verbal Span Backward	4.98 (1.260)	4.81 (1.331)	*F*<1, *p*=0.518
List Learning	31.10 (11.902)	36.09 (13.182)	3.880, *p*=0.052
Immediate Face-Name Recognition	18.17 (1.90)	17.71 (2.657)	1.019, *p*=0.315
Delayed Face-Name Recognition	9.51 (0.933)	9.35 (1.041)	*F*<1, *p*=0.431
Trail Making A	21.325 (5.677)	32.363 (9.021)	54.499***
Trail Making B	42.368 (11.454)	69.341 (20.336)	68.120***
Letter Fluency	48.00 (13.51)	48.43 (13.65)	*F*<1, *p*=0.875
Corsi Block Span	6.11 (0.954)	4.66 (0.815)	66.240***

***p<0.001, **p<0.01

3.3: Reaction Time Results

While we did not have any specific hypotheses related to reaction time differences between conditions, we analyzed the data to investigate the possibility of condition differences in reaction times or a speed-accuracy tradeoff. We conducted a 2 (age group) x 6 (condition) repeated measures ANOVA. There was no main effect of condition [$F(5, 505)<1$]. There was a main effect of age group [$F(1, 101)=16.571$, $p<0.001$] indicating that young adults were significantly faster on the memory task than older adults. There was not however, a condition x age group interaction [$F(5, 505)<1$] indicating young adults were significantly faster than older adults in all conditions. This effect is not surprising given general age-related slowing on tasks of this nature.

3.4: Performance Data Results

Recognition memory accuracy for the paired-associates task was calculated as the probability of hits minus the probability of false alarms. A hit was defined as an “intact” response to an intact pair. A false alarm was defined as an “intact” response to a rearranged pair. Additional response categories include misses, which are “rearranged” responses to intact pairs, and correct rejections, which are “rearranged” responses to rearranged pairs. For the purpose of this analysis, we considered hit minus false alarm rates (i.e. Pr estimates) because this measure controls for response bias i.e. someone could have high rates of hits because they always answer “intact” but they would also have high false alarm rates as well. We did not consider other response types because participants may use, for example, a recall to reject strategy when responding to rearranged pairs. By using hit minus false alarm rates, we were confident we were only measuring one process, that is, recollection-based recognition for the encoded pair. Additionally, all trials in which the participant did not respond (null trials), responded with an RT of 200ms or less, or responded more than once per trial were removed prior to analysis. All performance data for young and older adults, along with reaction time data, for each condition are presented in Tables 5. Refer to Figure 2 for corrected recognition accuracy data for each condition for both age groups.

Table 5. Mean response proportions for young and older adults including reaction time data, standard deviations in parentheses

	Young Adults		Old Adults	
Condition				
High Arousal, Different Song				
Hit	0.77 (0.14)	1604 (261)	0.73 (0.16)	1760 (321)
Miss	0.23 (0.14)	1882 (517)	0.27 (0.16)	2083 (535)
Correct Rejection	0.68 (0.21)	1709 (367)	0.50 (0.21)	2142 (467)
False Alarm	0.32 (0.21)	1823 (685)	0.50 (0.21)	1960 (421)
High Arousal, Same Song				
Hit	0.80 (0.13)	1574 (247)	0.72 (0.16)	1760 (281)
Miss	0.20 (0.13)	1890 (560)	0.28 (0.16)	2189 (517)
Correct Rejection	0.68 (0.23)	1837 (294)	0.54 (0.24)	2087 (599)
False Alarm	0.32 (0.23)	1741 (653)	0.46 (0.24)	1947 (433)
Low Arousal, Different Song				
Hit	0.79 (0.15)	1594 (221)	0.74 (0.15)	1792 (282)
Miss	0.21 (0.15)	1898 (625)	0.26 (0.15)	2190 (412)
Correct Rejection	0.69 (0.20)	1800 (251)	0.58 (0.23)	2059 (412)
False Alarm	0.31 (0.20)	1926 (450)	0.42 (0.23)	1998 (432)
Low Arousal, Same Song				
Hit	0.81 (0.11)	1582 (240)	0.73 (0.16)	1787 (326)
Miss	0.19 (0.11)	1889 (386)	0.27 (0.16)	2068 (541)
Correct Rejection	0.70 (0.20)	1832 (329)	0.51 (0.22)	2129 (555)
False Alarm	0.30 (0.20)	1955 (605)	0.49 (0.22)	1990 (420)
Musical Rain				
Hit	0.79 (0.12)	1547 (204)	0.74 (0.14)	1782 (325)
Miss	0.21 (0.12)	1805 (477)	0.26 (0.14)	2187 (403)
Correct Rejection	0.67 (0.23)	1795 (314)	0.53 (0.23)	1979 (574)
False Alarm	0.33 (0.23)	1831 (584)	0.47 (0.23)	1948 (439)
Silence				
Hit	0.80 (0.13)	1574 (247)	0.78 (0.14)	1754 (314)
Miss	0.20 (0.13)	2020 (524)	0.22 (0.14)	2213 (479)
Correct Rejection	0.69 (0.20)	1753 (407)	0.59 (0.20)	2152 (325)
False Alarm	0.31 (0.20)	1962 (505)	0.41 (0.20)	2069 (460)

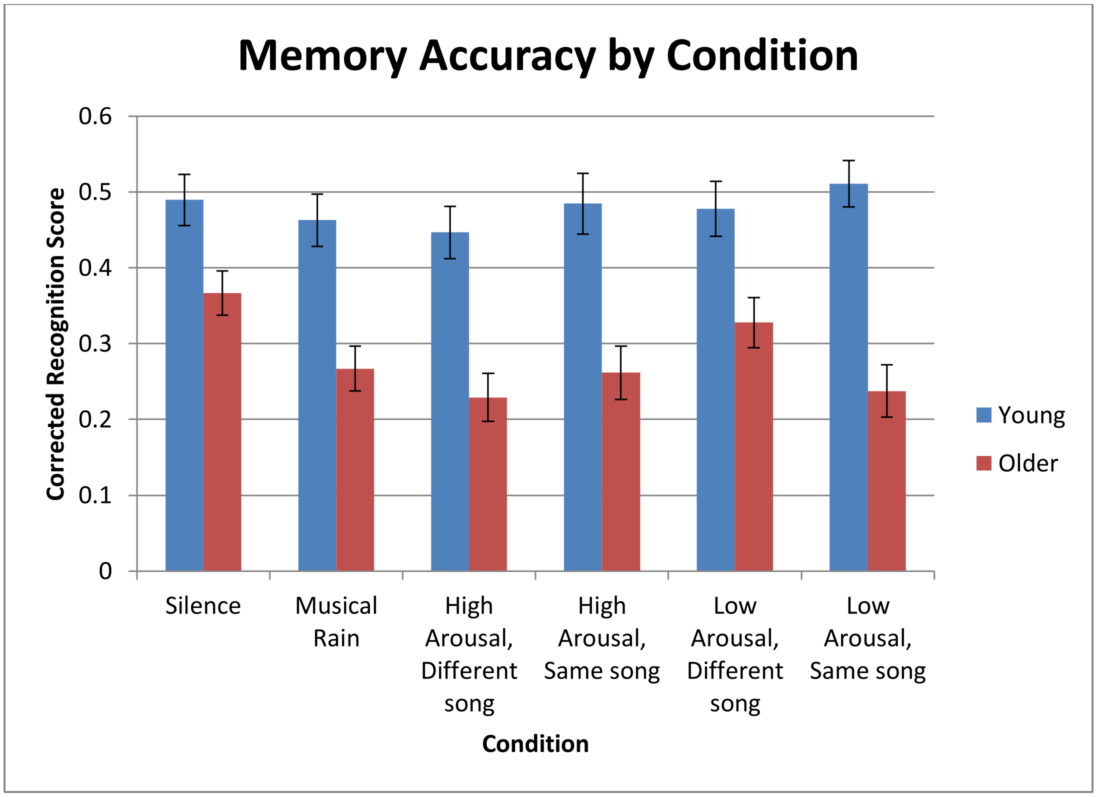

Figure 2. Corrected associative recognition (Pr) estimates as a function of condition for both young and older adults.

We conducted an omnibus repeated-measures 2 (age group) x 6 (condition) ANOVA on memory accuracy (Pr) estimates with age group as a between-subjects factor with two levels (young, older) and task condition as a within-subjects dependent factor with six levels (high arousal & different song [HAds], high arousal & same song [HAss], low arousal & different song [LAds], low arousal & same song [LAss], silence, and musical rain). There was a main effect of condition [F(4.839, 488.788)=4.962, p<0.001, partial η^2=0.047, observed d=0.981]. There was also a significant main effect of age group [F(1,101)=23.598, p<0.001, partial η^2=0.189, observed d=0.998]. The condition x age group interaction as also significant [F(4.839,488.788)=3.669, p<0.01, partial

η^2=0.035, observed *d*=0.923] indicating the differences in memory performance scores between conditions differed as function of age.

We followed up the omnibus ANOVA with the same repeated-measures ANOVA as above for each separate age group. For the young adults, there was not a main effect of condition [$F(4.835, 251.396)=1.282$, $p=0.273$, partial η^2=0.024, observed *d*=0.443] indicating there was no difference in memory accuracy between conditions. For older adults, however, there was a main effect of condition [$F(4.835, 237.787)=7.123$, $p<0.001$, partial η^2=0.127, observed *d*=0.998].

We followed up the main effect of condition for older adults with a series of paired-samples t-tests comparing the experimental conditions, as well as the musical rain control condition, to the silent control condition. After Bonferroni correcting based on conducting 5 t-tests, results indicated that older adults' memory performance in the musical rain, High Arousal (different song), High Arousal (same song), and Low Arousal (same song) conditions were all significantly lower than performance in the silent control condition. Performance in the Low Arousal (different song) condition was not significantly different from performance in the silent condition. See Table 6 for statistics related to this series of t-tests.

Table 6 Paired t-test statistics comparing memory performance in experimental and musical rain condition to performance in silent condition for older adults

Pair	Mean Difference (SD)	*t* (49)
Silent – HAds	0.138 (0.209)	4.663***
Silent – HAss	0.105 (0.205)	3.624***

Table 6 (continued).

Silent – LAds	0.039 (0.195)	1.411
Silent – LAss	0.129 (0.235)	3.883***
Silent – MR	0.100 (0.203)	3.461***

***p<0.001

3.5: Arousal and Context Results

We outlined two primary hypotheses related to the effects of arousal and context specificity:

1. Memory performance in high arousal conditions would be better than performance in low arousal conditions.
2. Memory performance in conditions wherein the same music is played during study and test will be better than conditions in which different music is played during study and test.

For this analysis, we corrected Pr measures for each condition for baseline, such that the Pr estimate for the silent control condition will be subtracted from each Pr estimate for the other conditions. Figure 3 presents scores in each condition for the different age groups as relative change from baseline.

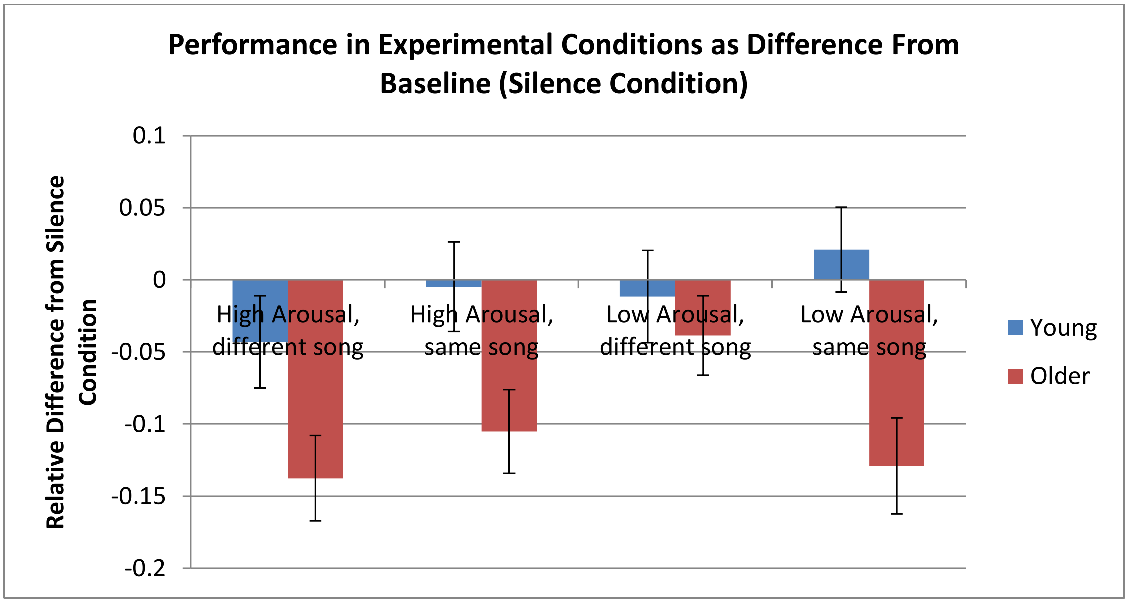

Figure 3. Performance data for young and older adults represented as relative to silent baseline condition.

We conducted an 2 x 2 x 2 repeated measures ANOVA on memory performance data (Pr) to test the main effects of arousal (High, Low) and context specificity (Same song, Different song), as well as their possible interactions with age group. There was a significant main effect of arousal [$F(1, 101)=7.726$, $p<0.01$, partial $\eta^2=0.071$, observed $d=0.786$]. There was also a significant main effect of age group [$F(1, 0.889)=6.354$, $p<0.05$, partial $\eta^2=0.059$, observed $d=0.704$]. There were significant interactions between context specificity and age group [$F(1, 101)=5.286$, $p<0.05$, partial $\eta^2=0.050$, observed $d=0.624$], arousal and context specificity [$F(1, 101)=5.932$, $p<0.05$, partial $\eta^2=0.055$, observed $d=0.674$] and a three-way interaction of age group, arousal, and context specificity emerged [$F(1, 101)=4.998$, $p<0.05$, partial $\eta^2=0.047$, observed $d=0.600$]. The main effect of context specificity was not significant [$F(1, 101)<1$, partial $\eta^2=0.001$, observed d=0.056], nor was the arousal by age group interaction [F(1, 101<1, partial η^2=0.001, observed d=0.065].

We followed up this significant omnibus ANOVA with separate 2 (arousal) x 2 (context specificity) repeated measures ANOVAs for each age group. For young adults, the main effect of context specificity was not significant [F(1,52)=3.033, p=0.088, partial η^2=0.055, observed d=0.401], however there was a trend for a main effect of arousal [F(1, 52)=3.731, p=0.059, partial η^2=0.067, observed d=0.474]. The interaction of arousal and context specificity for young adults was not significant [F(1,52)<1, partial η^2<0.001, observed d=0.052]. For older adults, the main effect of context specificity was not significant [F(1,49)=2.290, p=0.137, partial η^2<0.045, observed d=0.317], however, there was a trend for a main effect of arousal [F(1, 49)=3.981, p=0.052, partial η^2=0.075, observed d=0.498]. The interaction of arousal and context specificity was significant for older adults [F(1,49)=12.259, p<0.001, partial η^2=0.200, observed d=0.929].

Though the main effect was not significant, we decided to follow up the near-significant trend of the arousal main effect with a series of t-tests on condition difference for young adults. See Table 7 for complete pairwise comparison results for this data. The only pairwise difference that emerged was a significant difference between High Arousal, different song and Low Arousal, same song such that performance in the Low Arousal, same song condition was significantly higher than in the High Arousal, different song [$t(52)$=-2.350, p<0.05]. This result suggests that higher arousal music may distract young adults from task performance, though this was not significant across all high to low condition comparisons. We also see here some evidence of the predicted effect of context specificity in which playing the same song during study and test improved memory performance.

Table 7 Pairwise comparison data for young adults following up near-significant main effect of arousal

Pair	Mean Difference (SD)	*t* (52)
HAds – HAss	-0.038 (0.209)	-1.326
HAds – LAds	-0.031 (0.172)	-1.321
HAds – LAss	-0.064 (0.198)	-2.350*
HAss – LAds	0.007 (0.166)	0.295
HAss – LAss	-0.026 (0.184)	-1.030
LAds - LAss	-0.033 (0.201)	-1.188

*$p<0.05$

We explored the significant interaction of arousal and context specificity for older adults with a series of paired samples t-tests to which condition differences were driving this interaction. See Table 8 for complete pairwise comparison results for this data. Older adults' performance for the low arousal different song condition was significantly higher than in any of the other experimental conditions. This suggests that low arousal music is less detrimental to older adults' memory performance, and that context specificity may not have as great of an effect. This result is also similar to the result for young adults wherein memory performance in the Low Arousal, same song condition was higher than the High Arousal, different song condition.

Table 8 Pairwise comparison data for older adults following up significant interaction of arousal and context

Pair	Mean Difference (SD)	*t*	df
HAds – HAss	-0.033 (0.193)	-1.193	49
HAds – LAds	-0.099 (0.155)	-4.497***	49
HAds – LAss	-0.009 (0.201)	-0.300	49
HAss – LAds	-0.066 (0.177)	-2.648*	49
HAss – LAss	0.024 (0.204)	0.833	49
LAds - LAss	0.090 (0.173)	3.689***	49

*$p<0.05$, ***$p<0.001$

3.6: Subjective Ratings Results

We conducted an exploratory analysis on data collected following the memory test for each condition in which participants reported whether they found the music playing during the memory task was helpful or distracting to their performance. Participants responded to these questions on a 1 to 5 categorical scale, with 1 representing the extreme negative for the helpful question (i.e. “Very unhelpful”) and 5 representing negative for the distracting question (i.e. “Very distracting”). For example, if a participant thought the music in a condition was somewhat helpful and not at all distracting, they might respond 4 to the helpful question and 1 to the distracting question. Table 9 presents all of the mean responses for young and older adults. As can be seen in the table, neither group reported any of the conditions to be especially helpful, with all means below 4, with 5 being “Very helpful.” Given that there were two blocks for each condition, we averaged responses for each question across blocks to create a mean helpfulness and mean distracting rating for each condition. Subsequent analyses are based on this mean rating. 4 young adult participants did not complete these ratings, therefore this analysis is based on a sample of 50 young and 50 older adults.

Table 9. Mean Helpfulness and Distracting rating for young and older adults

Condition	YA Mean Helpful (SD)	YA Mean Distracting (SD)	OA Mean Helpful (SD)	OA Mean Distracting (SD)
HAds	2.7245 (0.936)	2.459 (0.956)	2.270 (0.938)	2.55 (1.121)
HAss	2.816 (0.808)	2.071 (0.777)	2.490 (0.860)	2.420 (1.002)
LAds	2.837 (0.780)	2.204 (0.901)	2.540 (0.891)	2.190 (1.044)
LAss	2.857 (0.743	2.184 (0.876)	2.680 (0.807)	2.080 (0.865)
Musical Rain	1.745 (0.778)	3.470 (1.129)	1.810 (0.721)	3.210 (1.183)
Silence	2.901 (0.546)	1.233 (0.422)	2.850 (1.093)	1.360 (0.887)

To explore between condition differences for both age groups, we started by conducting a 2 (age group) x 6 (condition) repeated measures ANOVA separately for helpfulness and distracting ratings. The omnibus ANOVA for helpfulness ratings revealed a significant main effect of condition [$F(5,93)=24.658$, $p<0.001$, observed $d=1.000$]. There was also a significant between-subjects main effect of age group [$F(1,97)=6.422$, $p<0.05$] indicating young adults tended to have higher helpfulness ratings than older adults. There was, however, no interaction between condition and age group, so ratings of helpfulness for young and older adults differed similarly across conditions. The repeated measures ANOVA for distracting ratings also revealed a main effect of condition [$F(5,93)=46.889$, $p<0.001$, observed $d=1.000$]. In the case of the distracting ratings, unlike the helpfulness ratings, there was not a main effect of age group, nor was there a condition x age group interaction.

We followed up the significant main effects of condition and age on helpfulness ratings with separate repeated measures ANOVAs for young and older adults to determine how young and older adults rated the music in separate condition as differentially helpful. For young adults, there was a significant main effect of condition [$F(5, 44)=14.808$, $p<0.001$, observed $d=1.000$]. The ANOVA for older adults also revealed a significant main effect of condition on helpfulness ratings [$F(5, 45)=10.726$, $p<0.001$, observed $d=1.000$].

We followed up the significant main effects of condition with separate paired t-test analyses on the ratings of helpfulness for both young and older adults. Table 10 presents complete paired t-test information for both young and older adults. After Bonferroni correcting for 15 t-tests, significant differences emerged for both young and older adults. The pattern of significant pairwise differences revealed that young adults rated the music in all experimental conditions, as well as the silence, as significantly more helpful than the musical rain noise. A similar pattern emerged for older adults in which they rated the all of the music, as well as the silence, as significantly more helpful than the musical rain noise. Additionally, older adults rated the music in the Low Arousal conditions as significantly more helpful than the music in the High Arousal, different song condition.

Table 10. Paired t-test information for helpfulness ratings for both young and older adults

	Young		Older	
Pair	Mean Difference (SD)	*t* (48)	Mean Difference (SD)	*t*(49)
HAds-HAss	-0.09 (0.78)	-0.83	-0.22 (0.78)	-2.00
HAds-LAds	-0.11 (1.23)	-0.64	-0.27 (0.87)	-2.19*
HAds-LAss	-0.13 (1.12)	-0.83	-0.41 (0.82)	-3.54***
HAds-MR	0.98 (1.21)	5.66***	0.46 (1.07)	3.03**
HAds-Silence	-0.18 (1.04)	-1.24	-0.58 (1.54)	-2.66*
HAss-LAds	-0.02 (1.09)	-0.13	-0.05 (0.74)	-0.48
HAss-LAss	-0.04 (1.03)	-0.28	-0.19 (0.76)	-1.76
HAss-MR	1.07 (1.16)	6.50***	0.68 (0.95)	5.08***
HAss-Silence	-0.09 (0.94)	-0.69	-0.36 (1.39)	-1.83
LAds-LAss	-0.02 (0.84)	-0.17	-0.14 (0.47)	-2.09*
LAds-MR	1.09 (1.07)	7.12***	0.73 (0.99)	5.21***
LAds-Silence	-0.07 (1.00)	-0.50	-0.31 (1.26)	-1.74
LAss-MR	1.11 (1.03)	7.58***	0.87 (0.92)	6.70***
LAss-Silence	-0.05 (1.00)	-0.36	-0.17 (1.12)	-1.07
MR-Silence	-1.16 (1.03)	-7.92***	-1.04 (1.33)	-5.54***

***$p<0.001$, **$p<0.01$, *$p<0.05$

Given that there was not a main effect of age group, nor an interaction of condition and age group, on distracting ratings, we collapsed rating across age groups to investigate differences between conditions. See Table 11 for paired t-test results, as well as distracting ratings based on our entire sample. This analysis resulted in an interesting pattern of differences. The first is that all participants rated the music in all conditions, as well as the silence, as less distracting that the music rain noise. Participants' felt, however, that the music was more distracting than silence. Secondly, participants rated the music in the High Arousal, different song condition as more distracting than both Low Arousal conditions and the Silence condition. Participants did not, however, rate music in the High Arousal, same song condition as more distracting than the low arousal conditions. These ratings tell us that while both young and old participants find music more helpful than general noise (music rain), they find it more distracting than silence.

Table 11. Paired sample t-test statistics for distracting ratings based on the entire sample of young and old participants

Pair	Mean Difference (SD)	*t* (98)
HAds-HAss	0.258 (0.882)	2.907
HAds-LAds	0.308 (1.022)	2.999*
HAds-LAss	0.374 (1.006)	3.697*
HAds-MR	-0.833 (1.374)	-6.035*
HAds-Silence	1.208 (1.200)	10.013*
HAss-LAds	0.051 (0.899)	0.559
HAss-LAss	0.116 (0.900)	1.248
HAss-MR	-1.091 (1.320)	-8.225*
HAss-Silence	0.949 (1.080)	8.750*
LAds-LAss	0.066 (0.903)	0.724
LAds-MR	-1.141 (1.392)	-8.156*
LAds-Silence	0.899 (1.069)	8.368*
LAss-MR	-1.207 (1.331)	-9.026*
LAss-Silence	0.833 (0.961)	8.628*

Table 11 (continued).

MR-Silence	2.040 (1.362)	14.904*

*p<0.003, criterion after Bonferroni correction

CHAPTER 4

DISCUSSION

We conducted a study to explore the concurrent effects of the arousal and context specificity of music on associative recognition memory performance for young and older adults. We hoped to show that music could be used as a memory aid to support associative recognition memory performance for both young and older adults, in particular for older adults who have deficits in recognition memory accuracy. This study extended upon current literature by comparing young and older adults' performance while listening to music and completing a real-world task, and included a representative sample of normally aging older adults rather than a patient population. We expected to see differences in task performance based on our two primary hypotheses. First, that arousal would influence memory performance by enhancing cognition in highly arousing conditions. Second, that manipulating the music context would influence performance such that when the musical context was the same between study and test, memory performance would be better than when the musical context was different. With respect to the
interaction of these factors, as well as their interaction with age, there is insufficient extant evidence in the literature to make specific predictions, however we hoped to show that the deficits older adults tend to show in the paired associate memory task we used may be ameliorated with the use of music.

In this study, we found that, in general, any background noise impaired older adults' performance relative to silence, whereas young adults were not impaired. While we hoped to show that arousing music would enhance memory performance, we also acknowledged the possibility that arousing music would distract older adults from the memory task, which was indeed what we saw, particularly for older adults. There was also no effect of context on memory performance, suggesting that reinstating memory context via background music was not particularly beneficial for memory performance. We did find that, relative to silence, although memory performance in most experimental conditions was impaired, older adults were not impaired in the Low arousal, different song condition. While this result is inconsistent with our predictions, it is not entirely inconsistent with extant literature. Greene et al. (2010) found that young adults' performance in a memory task was better when a low arousal, negative mood was induced. In our case, the fact that this condition did not impair performance despite a different song being played between study and test is somewhat difficult to explain. Although the finding that a low arousing background has little to no impact on memory performance in the old fits with the idea that more arousing distractors are more likely to draw attention away from the task of interest, this effect was only observed for one of the low arousal conditions. One possibility that we are currently exploring is that the particular songs, rather than the arousal and context categories per se, contribute to the results. That is, if we selected different songs from the rock genre or a different genre altogether, would the results be replicated? Nonetheless, this result indicates more research into the effects of arousal is necessary, but that for older adults, some music may be helpful, but highly arousing music is clearly not.

Participants also rated how helpful and distracting they felt the music, musical rain sound, and silence was to their memory performance. Both young and older adults felt that listening to music was more helpful to their performance than listening to noise (musical rain) but that music was not as helpful as silence. Older adults' performance substantiated these ratings since older adults' performance relative to silence was lower in all but one of the music listening conditions. Participants also rated music as more distracting than silence, which was also reflected in older adults' performance. Young adults also rated music as less helpful and more distracting than silence, but this was not reflected in their memory performance given that they did not perform differently from silence in the music conditions. It may be that older adults are more aware of their specific deficits when they are being distracted by background information, aware that they have some difficulty overcoming this distraction. It seems as though young adults are also aware that silence is the most helpful for their performance, but that they are able to inhibit processing the distracting musical stimuli, allowing them to remember information equally even if they are distracted. In this study, participants felt that music in the High Arousal, different song condition was the most distracting. Participants did not rate music in the High Arousal, same song condition as more distracting than other conditions and while this was not reflected in their memory performance, they might have felt the consistency of the song during study and test made music in the condition no more distracting than the music in other conditions.

We hoped to show that music could be used to support associative memory performance for older adults. Such a finding would be consistent with literature suggesting arousal can benefit cognition (e.g. Roth et al., 2008; Standing et al, 2008). Our results, however, show that listening to music impaired memory in the old. These results are consistent with work by Lynn Hasher and colleagues who proposed an inhibition deficit hypothesis in aging, in which they argue that older adults have difficulty inhibiting processing of extraneous information to the impairment of simultaneous task performance (Hasher & Zacks, 1988). In numerous experiments, they showed that older adults have difficulty inhibiting processing task-irrelevant stimuli, so their memory is impaired relative to young adults (e.g. Hasher, Stolzfus, Zacks, &Rypma, 1991; Ryan, Leung, Turk-Browne, & Hasher, 2007). In our study, older adults generally had difficulty inhibiting processing the music (or musical rain noise), so their memory performance tended to suffer. Hasher et al. (1988) argued that this process is two-fold, and that inhibitory failure occurs at both encoding and retrieval. During encoding, inhibitory mechanisms allow participants to attend to task-relevant stimuli rather than task-irrelevant stimuli. During retrieval, these inhibitory mechanisms allow participants to narrow attention to relevant memory searches. In the current study, we played music during both encoding and retrieval, and the failure of older adults' inhibitory mechanisms likely impaired both processes, distracting older adults and resulting in their generally poorer performance during those conditions.

Another possible reason for older adults' memory performance impairment during music listening conditions is that we used a sample of normally aging older adults. Much of the work to date using music to support memory for older adults has been done with patient populations. It may be that a normally aging sample of older adults does not benefit from using music as a memory aid similarly to older adults with memory impairments such as dementia. This is supported by the study by Simmons-Stern et al. (2010) who showed that healthy older adults did not remember significantly more children's song lyrics when they were paired with music versus simply spoken as the AD older adults patients did. More importantly, studies with patient populations generally use music as either the to-be-remembered material or as retrieval cues for long-term autobiographical memories. Both cases are clearly different from the current study wherein we did not ask participants to remember the music, and the music was not supporting retrieval of self-relevant information from long term memory. These mutually exclusive factors may explain why music did not improve memory performance for older adults in our study.

Our manipulation of the environmental context during study and test was not successful. We hoped to show that playing the same music during study and test would improve recognition memory accuracy above playing different songs during study and test. Such a finding would be consistent with the encoding specificity principle which states that memory will be enhanced when the retrieval context is similar to the encoding context. This, however, was not the case in our study wherein the change in context generally neither impaired nor improved memory performance in either age group. There are a couple possible reasons for the failure of this manipulation. One is that the ordering of the face-name pairs was not the same between study and test. Maintaining the exact ordering and presentation would not have been possible for the paired associates task in which some of the pair are rearranged. Furthermore, for the intact pairs, we swapped the position of the name and the face between study and test in order to reduce the probability of unitization of the pairs. Unitized information can be recognized on the basis of familiarity (Diana, et al., 2008). Given that familiarity, but not recollection may be spared in aging (Yonelinas, 2002), we created a memory that would more likely be based on recollection and likely reveal age-group differences in memory performance. The exact context of encoding could therefore not be replicated during retrieval.

There is some evidence suggesting we may have seen context effects in our study (Standing, et al., 2008). They found that paired associates recall of words was better when the same song was played during study and test than when a different song was played between study and test. There are, however, several differences in methodology between that study and ours. Namely, in the study by Standing and colleagues, background music was played in an adjoining room with an open door rather than via headphones as used in our study. Playing music through headphones may have resulted in the music being so distracting due to its auditory proximity (i.e. directly in the ears rather than distinctly in the background) that it distracted participants to a degree that any context effects were eliminated. Additionally, much of the encoding specificity literature (e.g. Dalton, 1993) emphasizes the context effects in memory have to do with the similarity between encoding and retrieval of the features of the stimuli, e.g. the font a word is presented in or the color of an image. Some of the features of the face-name stimuli, specifically the left-rightness of the names and faces, was intentionally changed to prevent ceiling effects on performance for young adults. Piloting in our lab has shown near-ceiling performance in similar tasks when the left-rightness of the pairs is not swapped. In this study, manipulating the order and presentation of the stimuli, as well as presenting music via headphones, may explain why our context specificity manipulation failed.

Given the results of this study, there are some lingering questions that remain that may be answered by a study using a task with even higher ecological validity. For example, an interesting experiment would be to use a real-world task environment, such as a driving simulator. During the study, young and older adults anecdotally reported listening to music while driving. In a complex task such as driving, older adults who listen to music while driving may fail to inhibit processing musical information, interfering with their ability to process the environmental information. Such interference could easily result in safety concerns for older adult drivers.

This study was an exploratory first step toward investigating the question of how we might use music to support memory for young and healthy aging older adults in the absence of a body of literature that we could use to make clear predictions about the pattern of results we should expect. This study showed that young and older adults perform differently on a memory task in the presence of background music, despite finding the music similarly distracting. While we did not show support for using music to support memory for either young or older adults, we did find that music differentially impairs young and older adults, and that this this impairment may be associated with the arousing nature of the music.

REFERENCES

Baur, B., Uttner, I., Ilmberger, J., Fesl, G., & Mal, N. (2000). Music memory provides access to verbal knowledge in a patient with global amnesia. *Neurocase, 6*, 415-421.

Benton, A. L., Hamsher, S. K., & Sivan, A. B. (1983). *Multilingual aplasia examination*. Iowa City: AJA Associates.

Chabris, C. F. (1999). Prelude or requiem for the "Mozart effect"? *Nature, 400*, 826-827.

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain*. Unpublished doctoral dissertation. McGill University.

Dalton, P. (1993). The role of stimulus-familiarity in context-dependent recognition. *Memory & Cognition*, *21*(2), 223-234.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2008). The effects of unitization on familiarity-based source memory: Testing a behavioral prediction derived from neuroimaging data. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 730-740.

Greene, C. M., Bahri, P., & Soto, D. (2010). Interplay between affect and arousal in recognition memory. *PLoS ONE, 5*(7). doi: 10.1371/journal.pone.0011739

Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Rypma, B. (1991). Age and Inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 163-169.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 22, pp. 193-225). New York: Academic Press.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* Gainesville, FL.: University of Florida.

Liu, D. L. J., Graham, S., & Zorawski, M. (2008). Enhanced selective memory consolidation following post-learning pleasant and aversive arousal. *Neurobiology of Learning and Memory, 89*, 36-46.

Lord, T. R., & Garner, J. E. (1993). Effects of music on Alzheimer patients. *Perceptual and Motor Skills*, *76*(2), 451-455.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1170-1187.

Naveh-Benjamin, M., Guez, J., Kilb, A., & Reedy, S. (2004). The associative memory deficit of older adults: Further support using face-name associations. *Psychology and Aging*, *19*(3), 541-546.

Naveh-Benjamin, M., Hussain, Z., Guez, J., & Bar-On, M. (2003). Adult age differences in episodic memory: Further support for an associative-deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 826-837.

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging, 23*(1), 104-118.

Peretz, I., Champod, S., & Hyde, K. (2003). Varieties of music disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences, 999*, 58-75.

Purnell-Webb, P., & Speelman, C. P. (2008). Effects of music on memory for text. *Perceptual and Motor Skills, 106*(3), 927-957.

Rauscher, F. H., Shaw, G. L., & Ky, K., N. (1993). Music and spatial task performance. *Nature*, *365*, 611.

Rauscher, F. H., Shaw, G. L., & Ky, K., N. (1995). Listening to Mozart enhances spatial-temporal reasoning -- Towards a neurophysiological basis. *Neuroscience Letters, 185*(1), 44-47.

Reitan, R., & Wolfson, D. (1985). *The Halstead-Reitan Neuropsychological Test Battery: Therapy and clinical assessment*. Tucson, AZ: Neuropsychological Press.

Roth, E. A., & Smith, K. H. (2008). The Mozart effect: Evidence for the arousal hypothesis. *Perceptual and Motor Skills, 107*, 396-402.

Ryan, J. D., Leung, G., Turk-Browne, N. B., & Hasher, L. (2007). Assessment of age-related changes in inhibition and binding using eye movement monitoring. *Psychology and Aging, 22*(2), 239-250.

Salamé, P., & Baddeley, A. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology, 41A*(1), 107-122.

Simmons-Stern, N. R., Budson, A. E., & Ally, B. A. (2010). Music as a memory enhancer in patients with Alzheimer's Disease. *Neuropsychologia, 48*(10), 3164-3167.

Standing, L. G., Bobbitt, K. E., Boisvert, K. L., Dayholos, K. N., & Gagnon, A. M. (2008). People, clothing, music, and arousal as contextual and retrieval cues in verbal memory. *Perceptual and Motor Skills, 107*, 523-534.

Thompson, R. G., Moulin, C. J. A., Hayre, S., & Jones, R. W. (1995). Music enhances category fluency in healty older adults and Alzheimer's disease patients. *Experimental Aging Research, 31*, 91-99.

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology, 86*(2), 255-262.

Tulving, E., & Thomson, D. M. (1971). Retreival processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology, 87*(1), 116-124.

Uppenkamp, S., Johnsrude, I. S., Norris, D., Marslen-Wilson, W., & Patterson, R. D. (2006). Locating speech-specific processes in the human temporal cortex. *NeuroImage, 31*, 1284-1296.

Wallace, W. T. (1994). Memory for music: Effect of melody on recall of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1471-1485.

Watkins, M. J., Ho, E., & Tulving, E. (1976). Context effects in recognition memory for faces. *Journal of Verbal Learning and Verbal Behavior, 15*(5), 505-517.

Williams, J. (1991). *Memory assessment scales professional manual.* Odessa: Psychological Assessment Resources.

Yalch, R. F. (1991). Memory in a jingle jungle: Music as a mnemonic device in communicating advertising slogans. *Journal of Applied Psychology, 76*(2), 268-275.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(441-517).